Modelling and estimating modal share in European transport: a comparative analysis using international freight flow data

Russo, G.; Reggiani, A; Nijkamp, P.

2001

TI 2001-095/3
Tinbergen Institute Discussion Paper

Modelling and Estimating Modal Share in European Transport

Giovanni Russo[1,4]
Aura Reggiani[2]
Peter Nijkamp[3,4]

[1] *Department of Economics and Policy, Utrecht University,* [2] *Department of Economics, Bologna University,* [3] *Department of Regional Economics, Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam,* [4] *Tinbergen Institute*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Modelling and Estimating Modal Share in European Transport: a Comparative Analysis Using Interregional Freight Flow Data

Giovanni Russo
Utrecht University
Department of Economics and Policy
PO Box 80140
3508 TC Utrecht
The Netherlands
Tel +31-30-2531115
Email: g.russo@fss.uu.nl

Aura Reggiani
Bologna University
Department of Economics
Piazza Scaravilli 2
40126 Bologna
Italy

Peter Nijkamp
Free University
Department of Regional Economics
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

Utrecht, 9th September, 2001.

Abstract

International and interregional trade and transport are on the rise and hence, there is a clear need for reliable estimates of transport flows. However, the available databases and estimation methods are not yet satisfactory for analytical and predictive purposes. In this paper we explore the use of different statistical techniques in order to examine the spatial flow pattern of freight transport among competing transport modes.
Freight transport has specific peculiarities that are different from passenger transport. We argue that a logit model, the most commonly used technique in the empirical analysis of passenger flows, is not always appropriate for the analysis of freight flows, unless the interdependence between the decision making regarding the shipment of individual units of freight belonging to the same shipment is correctly modelled.
In the paper, we will focus on the analysis of aggregate freight transport flows of the type that may be generated from conventional spatial interaction models. In particular, adjusted estimation techniques alternative to the logit model will be employed to analyze the transport flows of two products, chemical products and foodstuffs, based on an European interregional data set.
We will conclude the paper by discussing various caveats encountered during the empirical analysis.

1. Introduction

As a result of globalisation, liberalisation and economic integration, a world-wide surge in trade and transport flows can be observed. At the same time, there is a growing awareness that externalities of various kinds (congestion, environmental stress, safety, etc.) may become severe stumbling blocks in the process of international or interregional trade. Hence, we observe an increasing interest in optimal capacity use of the available infrastructure, *inter alia* in the area of modal choice. This recognition has also prompted a renewed interest in spatial interaction modelling, particularly in the European context.

Generally speaking, a flow of goods between two regions takes place when the difference in the market clearing prices in the two regions is larger than the cost of transport, provided that the two regions are linked by an adequate infrastructure (see also Rietveld and Nijkamp 2000).

Accessibility plays clearly an important role in firms' locational decisions. *Ceteris paribus*, more accessible regions tend to attract more firms that will generate more competitive output and thus a higher level of economic activity. Therefore, the level of imports and exports will, in turn, tend to rise accordingly (De Dios Ortúzar and Willumsen 1990).

In order to move goods from the origin to the destination one needs to use carriers. Therefore, the structure of the carriers' market is also of critical importance (e.g. whether carriers collude or not). In a non-competitive carriers' market, prices are higher than the corresponding marginal costs. The amount of transport in a non-competitive market is, therefore, lower than the social optimum (Hurley and Petersen 1996a, 1996b). A too high price of transport will also have a feedback effect on firms' location and firms' inventory management decisions.

It is noteworthy that, in general, in the literature, the demand for freight transport has received much less attention than passenger demand. In fact, for a long time congestion in urban areas has been (and still is) prominent on the political agenda in many industrialised countries, and in this framework passenger movements play a much more important role than freight transport.

On the other hand, from the perspective of a unified European market, freight transport is likely to gain more importance, as spatial competition is likely to become an important source of competitive advantage.

Freight transport presents, on the one hand, features that make it similar to passenger movements; for example, price considerations affect the way the flow of both goods and passengers is distributed across transport modes in the same way. On the other hand, freight transport also has various intrinsic peculiarities that make it different from the standard model of passenger movements; for example, the existence of volume pricing implies that in freight transport the unit cost of transport interacts with the volume of the shipment. These two factors jointly determine the average cost of transport. It should thus be noticed that the market structure might have very important effects on the interaction between the volume of goods between two regions via a given mode of transport and the relative average cost of the shipment. For example, suppose that in a certain region only one carrier (a monopolist) serves one mode of transport (equivalently, more carriers may also collude and behave like in a shared monopoly). If the carriers apply price discrimination (of the second type) in the form of volume discounts, the decision to ship one additional tonne of freight along a given link depends on the number of tonnes of freight already scheduled to be shipped along that link. This phenomenon introduces interdependence between the tonnes of freight belonging to the same shipment. This, in turn, has implications for the empirical analysis of such data. In fact, statistical models that assume that the choice of the mode of transportation is made independently for each tonne (such as the logit model, also in its grouped data version) would fail with this respect.

Against the previous background, the aim of this paper is to offer an exploratory investigation of estimation techniques that may be used to analyse how flows of freight transport between origin-destination pairs in spatial interaction models are allocated to competing transport modes.

The paper is organized as follows. In section 2 we show the implications of the type of modelling chosen (micro or macro) for the empirical analysis. Section 3 describes the available data sets, and section 4 presents the results from the empirical analysis. Finally, section 5 contains some concluding remarks.

2. Statistical Analysis of Transport Flows

In the present paper we will offer a statistical analysis of freight flows from the perspective of spatial interaction models[1]. An important empirical question concerns the analysis of the process governing the allocation of freight flows on a certain link to competing transport modes. The allocation process may be influenced by such factors as the unit cost of transport via the competing transport modes, the difference in travelling time and/or the difference in the distance to be travelled when using either of the two transport modes[2].

Statistical methods can be used to analyse whether there is a systematic – possibly causal – relationship between the aforementioned variables and the quantity of freight flows shipped via two competing transport modes (in our case, road and rail).

At first glance, this problem resembles very much the problem of modal split in passenger transport. Clearly, there is a similarity in flow data, such as the total flow of passengers between any two locations, and the number of passengers travelling via each of the competing transport modes. But there are also methodological differences. The flow of passengers travelling via any of the competing transport modes is the result of a process of (individual) utility maximisation (Ben-Akiva and Lerman 1985). Individual travellers choose the preferred transport mode independently from each other (except in the case of congestion; see Emmerink 1996). The independence of individual decision-making implies that the observed flow can be considered as a compact representation of the underlying individual data. In the latter context, discrete choice models can be used to investigate the agents' decisions with regard the preferred transport modes (Ben-Akiva and Lerman 1985).

It is, given the similarity in flow data in freight and passenger transport, tempting to apply *mutatis mutandis* the same statistical techniques to the choice of the mode of transport in both cases. This would certainly be correct, if each tonne of freight transport could be considered as an independent individual unit, so that the choice to

[1] Transportation networks can be represented as Input–Output tables with origins arranged down the columns and destinations arranged along the rows (see e.g. Nijkamp and Reggiani 1989). Each cell in this origin–destination table represents the total flow of freight from a given origin to a given destination. Similarly, in the event that origin–destination pairs are linked by more routes (involving different transport modes, e.g. roads and railways) separate origin–destination tables for each of the transport modes can be constructed.

[2] This list pretends by no means to be an exhaustive list of the determinants of the decision to choose a particular means of transport.

ship this tonne via either of the transport modes could be made independently for each tonne.

The key problem now lies in the fact that, in freight transport, a certain flow (i.e., volume of tonnes) of a given commodity does not necessarily correspond to n agents each independently choosing the preferred mode of transport to ship their own tonne. In fact, the number of decision-makers is normally rather limited, as the freight market has rather oligopolistic features (see also NCHRP 1997). Furthermore, the existence of quantity discounts (volume pricing) implies that the decision to ship one tonne of commodities via a certain mode of transport is not independent from the decision on the remaining tonnes of commodities belonging to the same shipment. In this way, the link between individual transport choice behaviour and aggregate flow outcomes for freight transport becomes non-linear. Therefore, stochastic utility models should be able to address the issue of the interdependency between tonnes belonging to the same shipment when the choice of transport mode is analyzed. In other words, the correlation between the error terms in the equations referring to the tonnes of freight in the same shipment is non-zero. This problem can be solved by considering the tonnes of freight belonging to a given shipment as repeated observations, as is done in Random Effect Panel Data analysis (Green 2000). To apply this technique, the number of shipments along a given link should be known. However, this is generally not the case when dealing with aggregate data from spatial interaction models (a case faced in the present analysis). It is, therefore, intriguing to examine how the same macro data used to calibrate spatial interaction models can be used to investigate the relationship between the average amount of freight shipped via a given transport mode and the average cost, travel time and distance associated with it (and with the competing transport modes). To this end, we deploy an adjusted regression method (see section 4). Before presenting these econometric experiments we will first present in section 3 the data sets used.

3. Description of the Data Sets

3.1. General

The data sets used in our analysis concern the flow of goods (foodstuffs and chemical products) between 108 regions belonging to 14 countries in the EU in the year 1986

(the regional classification can be found in Appendix 1; additional information on the data can be found in Buratto 1999). Consequently, there are, in principle, 11664 possible links. These regions are used for a spatial interaction modelling effort and do not entirely correspond to the standard European regional classification; therefore, available regional data (i.e., regional income and population as published by Eurostat) are – at least – difficult to match. Moreover, we neglect border impediments such as the different railway gauges between France and Spain and the different train lengths between various countries.

Our aim is, in fact, rather modest, as we want to show how these potentially interesting data can meaningfully be used in an empirical analysis. The data are also interesting because the dependent variable, viz., the share of the total freight flow on a given link shipped by road, presents different characteristics in the two data sets; these characteristics will bear on the statistical techniques used to extract the information contained in the data.

The data contain information on the flows of goods between each pair of regions; flows of goods within regions are not considered. Moreover, the data sets contain information on the total transport costs over the links, the distance, and the travel time between origins and destinations via different transport modes (road and rail). This information is separately available for two types of commodities: foodstuffs and chemical products. Additional information on the data can be found in Buratto (1999), Nijkamp and Reggiani (1998), and Reggiani (1998). The list of the variables used is presented below.

In the case of foodstuffs we have the following variables:

RC: the transport cost between any two regions by road (Euros/tonne).

TC: the transport cost between any two regions by rail (Euros/tonne).

RT: the travel time between any two regions by road (minutes).

TT: the travel time between any two regions by rail (minutes).

s: the share of the flow of foodstuffs between any two regions routed by road.

RELC: the relative transport cost defined as RC/TC.

RELC2: the square of the relative transport cost, defined as $(RELC)^2$.

RELT: the relative travel time, defined as RT/TT.

RELT2: the square of the relative travel time, defined as $(RELT)^2$.

In the case of chemical products the data contain information on the following variables:

RC: the transport cost between two regions by road (Euros/tonne).

TC: the transport cost between two regions by rail (Euros/tonne).

RD: the distance between two regions by road (Kms).

TD: the distance between two regions by rail (Kms).

*s*1: the share of the flow of chemical products between two regions routed by road.

Finally, the mean and standard deviation of the variables used are presented in Table 1.

Table 1 : Descriptive statistics of the variables used.

Foodstuffs			Chemical Products		
# observations: 3439			# observations: 1731		
	Mean	Std. Dev.		Mean	Std. Dev.
RC	79.61	35.18	RC	78.52	34.29
TC	84.61	27.07	TC	90.24	31.74
RT	106.21	687.87	RD	1039.46	603.99
TT	1308.57	777.44	TD	1086.26	631.11
s	0.95	0.22	*s*1	0.82	0.19

An unbalanced distribution of the number of observations across modes of transport (i.e., more than 80% of the flow of freight is routed by road) can impair the predictive ability of any statistical model (and discrete choice models in particular, Cramer 1996). It is evident from Table 1 that this is the case in both data sets. The number of observations refers to the number of origin - destination pairs with non-zero flows on at least one of the two modes of transport (road or rail). Moreover, foodstuffs display a significantly shorter travel time by road than by rail, while chemical products travel similar distances by road and rail.

Some specific statistical proprieties of the data sets will be discussed in subsection 3.2 and 3.3.

3.2. Foodstuffs

The data set contains 3439 observations on foodstuffs flows between EU regions (origin-destination pairs). This means essentially that a non-zero flow of foodstuffs

has been observed on 29.5 % of all possible links. The average foodstuffs flow is 39817 tonnes of which 37760 tonnes are routed by road, and the remaining 2058 tonnes by rail. On average, 95% of the freight is routed by road. However, these figures are somewhat misleading. In fact, the median flow[3] of foodstuffs is as low as 1168 tonnes. The median flow of foodstuffs by road and rail is as low as 1068 tonnes and 7 tonnes, respectively[4]. The large average flow is thus the result of the combination of very large flows on a relatively few links and very small flows on relatively many links (i.e., the distribution of flows is skewed). The flow of foodstuffs in 2.6% of the cases is entirely routed by rail, while in 38.6% of the cases it is entirely routed by road.

It is noteworthy that the explanatory variables included in the data set appear to be highly correlated; in fact, the correlation coefficient between the travel time by road and by rail is 0.98 (significant at 5%). Likewise, the correlation coefficient between the cost of shipping one tonne of foodstuffs by road and via the rail is 0.85 (significant at 5%). Given the very large share of the total flow of foodstuffs that is routed towards their destination by road, the very high correlation coefficient between these two flows (0.99, significant at 5%) comes as no surprise.

3.3. Chemical products

The data sets on chemical products contains 1731 observations on non-zero flows between European regions (origin-destination pairs)[5]. The average flow amounts to 39255 tonnes, of which 31898 tonnes are routed by road and the remaining 7356 tonnes by rail. If we look at the medians, we see that the median total flow between regions amounts to only 4375 tonnes, the median flow by road amounts to 3465 tonnes, and the median flow by rail is as low as 356 tonnes. The quite large average flows are thus the result of the combination of very large flows on a relatively few links and of very small flows on relatively many links (the distribution of flows is again very skewed). Since a consistent share of the total flow is routed by road (82%

[3] 50% of the flows consist of a number of tonnes smaller than or equal to the value of the median flow.
[4] The median flow of foodstuffs calculated without including the links with a zero flow amounts to 2341 tonnes. The median flow of foodstuffs via road and railway is 2046 tonnes and 73 tonnes, respectively.
[5] The number of observations in the two analyses may differ, because only those regions linked by non-zero flows are considered.

on average), the flow of chemical products by road between any two regions is highly correlated with the total flow (the correlation coefficient is 0.98, significant at 5%).

4. The Empirical Analysis

4.1. General

In this section we explore the distribution of the flow of foodstuffs and chemical products between the two competing transport modes, viz. the road and the railway. In both the foodstuffs and the chemical products cases the dependent variable is a share, viz. the share of the relevant good shipped by road. This variable is not obtained from individual surveys, but it is rather an aggregate figure (similar to those used in spatial interaction models). It might be tempting to treat this share as the outcome of the aggregation of independent individual choices. This bears a seemingly striking resemblance to the share of passengers travelling by car when the data are presented as grouped data. This parallel is correct, if the decision to ship each tonne of freight via either transport mode is made independently from the decision to ship other tonnes of freight. Because of the presence of quantity discounts and certain restrictions on the size of the containers, the independence assumption is untenable, at least for the tonnes belonging to the same shipment. In this case the estimation of discrete choice models from grouped data would be inappropriate. However, it would still be possible to retain the assumption of independence between shipments, and in the case that the tonnes of freight belonging to each shipment were known, one could estimate random effect discrete choice models from panel data. Should this information not be available, then one has to look for alternatives that are more agnostic about the structure of the errors. For example, these data could be considered as macro aggregates; in this case the only statistical requirement refers to the independence of the errors between origin – destination pairs. The ensuing empirical analyses are carried out taking into account the above-described constraints imposed by the data.

4.2. Empirical analysis of foodstuffs flows

The aim of our statistical analysis is to explain the allocation of foodstuffs flows between competing transport modes by some mode-specific background variables. The dependent variable in our analysis is the share (*s*) of the total flow of foodstuffs - on a given link - routed towards its destination by road. Furthermore, it must be considered that the range of variation of a share is the unit interval [0,1]. In a regression analysis the error term can vary on the entire real axis; therefore, an inconsistency arises between the range of variation of the dependent variable and that of the error term. This discrepancy may be accommodated by transforming the dependent variable into a new variable *g*, as follows: $g=\ln[s/(1-s)]$. This new variable *g* can vary from $-\infty$ to $+\infty$. This transformation is possible, if and only if the variable *s* (the share of freight routed by road) is different from 1 or 0, i.e., $0<s<1$. This is, however, not the case for the share of foodstuffs flow routed by road, as there are many zero entries in the spatial data matrix (the flow of freight on these links is routed entirely via one of the means of transport).

An alternative way to accommodate this discrepancy is to censor the distribution of the error term like in a two-limit tobit model (Maddala 1985). This model assumes the existence of a latent variable s^* which is not directly observable; instead, we observe only the realised variable *s* that is related to the latent variable s^*. The latent variable s^* is assumed to be a linear function of the independent variables and reads as follows:

$$s_i^* = \beta_0 + \beta_1 RC_i + \beta_2 RT_i + \beta_3 TC_i + \beta_4 TT_i + u_i \qquad [1]$$

where *u* is a normally distributed error term, assumed to be independently identically distributed (iid) across origin-destination pairs (denoted by the subscript *i*).

The latent variable s^* and the observed variable *s* are related in the following way:

$$s_i = 0 \quad if \quad s_i^* \le 0$$
$$s_i = s_i^* \quad if \quad 0 < s_i^* < 1$$
$$s_i = 1 \quad if \quad s_i^* \ge 1$$

The corresponding likelihood function (*L*) reads as follows:

$$
\begin{aligned}
L &= \prod_{s_i=0} \Phi_{1i} \prod_{s_i=s_i^*} \frac{1}{\sigma}\phi\left[\frac{s_i - R_i}{\sigma}\right] \prod_{s_i=1} (1-\Phi_{2i}) \\
R_i &= \beta_0 + \beta_1 RC_i + \beta_2 RT_i + \beta_3 TC_i + \beta_4 TT_i \\
\Phi_{1i} &= \Phi\left[\frac{-R_i}{\sigma}\right] \\
\Phi_{2i} &= \Phi\left[\frac{1-R_i}{\sigma}\right]
\end{aligned}
\qquad [2]
$$

where Φ is the cumulative distribution function of a normal probability function (ϕ) with average zero and variance σ^2. The effect of a unit change in one of the independent variables on the dependent variable (measured via the corresponding slope parameter β) has to be corrected for the effect that the change in the independent variable has on the probability that s is equal to zero or one (Maddala 1985). This adjustment may be represented as follows:

$$
E(s_i \mid 0 < s_i^* < 1) = R_i + \sigma \frac{\phi_{1i} - \phi_{2i}}{\Phi_{2i} - \Phi_{1i}} \qquad [3]
$$

where *E* represents the expectation operator. The average effect of all other variables that are not included in the model is captured by the constant term (β_0).

We can now estimate the two-limit tobit model presented in equation [2]. However, a link model specification test (Pregibon 1980) [6] signals that the model is mis-specified at the 5% confidence level[7]. One possible cause is most likely the presence of a high correlation between the regressors, especially with the variables referring to time. Consequently, we estimated a restricted model (with the following restriction: $\beta_2=\beta_4=0$). A likelihood ratio test (LR=6.7, significant at 5%, $\chi^2_{(2)}(5\%)=6$) rejects this restriction. The link model specification test again shows the presence of mis-specification. In order to cope with these unfavourable results, we tried alternative functional forms for the regressors, by using a logarithmic specification and adding quadratic terms, but the model remained poorly specified. In addition, we adjusted the model by specifying the share of foodstuffs routed by road as a function of the relative cost (per tonne) and the relative travel time of the two transport modes. The link test signals once more that the model is mis-specified. Fortunately, the link specification test did not reject the latter specification, when one includes quadratic

[6] The link model specification test is very similar to the Ramsey's RESET (Gujarati 1995) test in spirit. The dependent variable is regressed against powers of its predicted values (as obtained from the two-limit tobit model). One or more significant coefficients would signal mis-specification.

[7] All results are available from the authors upon request.

terms for the relative cost and travel time. Thus, this specification is apparently more satisfactory. The estimation results concerning this last model are presented in Table 2. The quadratic relationship between the relative road cost (relative to the railway cost, see section 3.1.) implies that, if the road transport cost is less than half the railway transport cost, then an increase of the relative road transport cost would actually increase the share of goods shipped by road. This may be due to some other comparative advantages (such as flexibility) that the road system may have in comparison to the railway system. At any rate, the relative cost factor when the road transport cost is more than half the railway transport cost overrules these factors. The higher the relative cost of road transport the lower the share of foodstuffs shipped by the road system, however. The explanatory power of the model is rather low (too many missing variables) to draw solid inference on the nature of the non-linearity in the relative costs and travel time variables.

Nonetheless, the flows of foodstuffs predicted by using the parameters estimated from this last two-limit tobit model highly correlate with the observed flows (the correlation coefficient is 0.99, significant at 5%).

Table 2: Estimation results of the two-limit tobit model (*s*= share of foodstuffs flow by road)

Dependent variable: *s*

			Observations	3439
	Coeff.	Std. Err.	LR Test	396
RELC	0.451	0.136	Pseudo R^2	0.122
RELC2	-0.451	0.068	90 left-censored observations at s<0	
RELT	-0.678	0.417	1327 right-censored observations at s>1	
RELT2	0.497	0.262	2022 uncensored observations	
Constant	1.207	0.059	$RELC2=(RELC)^2$	
σ	0.297	0.005	$RELT2=(RELT)^2$	

4.3. Empirical analysis of chemical product flows

In this second application, we concentrate on the interregional modal distribution of chemical products. Here, the dependent variable is the share of the flow of chemical products that is routed towards a given destination on the road system (*s*1). The dependent variable *s*1 is again a share, as in the foodstuffs application. But, unlike the case of foodstuffs, here the share of chemical products shipped by road – *s*1 – happens to be strictly included in the interval (0,1), $0<s1<1$.

In this case, the straightforward transformation g=ln[$s1$/(1-$s1$)] can be applied. This transformation is very similar to the one applied in the estimation of discrete choice models (or logit models) for grouped data. The only difference lies in the fact that the error term associated with this transformation needs not be heteroscedastic. Consequently, one can use a simple ordinary least squares (OLS) estimation procedure.

We set out to estimate the following regression:

$$g_i = \beta_0 + \beta_1 RC_i + \beta_2 RD_i + \beta_3 TC_i + \beta_4 TD_i + u_i \qquad [4]$$

where u is an error term satisfying the usual OLS assumptions (iid across origin-destination pairs, denoted by the subscript i). The estimated parameters are presented in Table 3. It appears that an increase in the road transportation cost and in the road travel distance decreases the share of chemical products shipped by road. On the contrary, an increase of the rail travel distance (the alternative mode of transport) tends to increase the share of chemical products shipped by road. These results are nicely in conformity with what is predicted from demand analysis: the demand of road transport via a given mode of transport depends positively on the price and distance of the concurrent mode of transport and negatively on its own price and distance.

After the above estimation, an application of the Ramsey RESET specification test (Gujarati 1995) demonstrates that the model is poorly specified. From an inspection of the residuals, the existence of an outlier can easily be identified. In order to neutralise its effect a dummy variable (D1) is included, where the value 1 corresponds to the observation responsible for the outlier, and zero otherwise. The inclusion of this dummy did improve the statistical fit of the model; in fact, the Ramsey RESET test no longer signals the presence of mis-specification (RESET=1.62, $F^{(5\%)}_{(12,1713)}$=1.75). The R^2 indicates that the model can explain only 4% of the variation in the dependent variable (though this is still significantly different from zero, as the F-test shows). Therefore, one may draw the conclusion that omitted factors (not included in the model) probably play a much more important role than the costs and the distance. Examples of such factors might be the physical geography of areas discouraging the use of the rail, or the logistic requirement caused by a specific network configuration. Because of the high correlation between the distance variables and the relative cost variables, we have also tested the following restriction: Ho: $\beta_2 = \beta_4 = 0$ against the

alternative hypotheses H1: $\beta_1 \neq 0$, $\beta_4 \neq 0$. The restriction is rejected by an F-test at a 5% confidence level (F=20.04, $F_{(2,1725)}^{5\%}=3$).

If the average cost is a proxy for the shipment price (i.e., p=RC), the price elasticity of the demand for freight transport via one mode of transport can be derived as follows:

$$\eta_{sp} = \frac{ds}{dp}\frac{p}{s1} = \beta_1 \left[\frac{p}{s1(1-s1)} \right] \qquad [5]$$

This expression, when evaluated at the sample average of p and s, returns an approximation of the average price elasticity of the demand of freight transport. In our case, we find that the price elasticity of the share of chemical products shipped by road is price-elastic (η_{sp}=-1.23). Consequently, a 10% increase in the average road cost (about 8 Euros per tonne) would decrease the share of chemical products shipped by road by 12.3% (i.e. 0.1, from 0.82 to 0.72)[8].

Despite some shortcomings in the simple specification of the model, it appears to successfully capture the behaviour of the flow of chemical products routed by road. The correlation between the real and the predicted flow of chemical products by road is even as high as 0.98, significant at a 5%.

Table 3: Results of the OLS model (g= (transformation of) the share of chemical products flow by road)

Dependent variable: g

	Coeff.	Std. Err.	Observations	1731
RC	-0.018	0.004	R^2	0.04
RD (/10)	-0.018	0.005	F-test	15.43
TD (/10)	0.026	0.005	F(5,1725) at 5%	2.21
TC	0.001	0.002		
D1	9.519	1.655		
Constant	2.581	0.014	g=ln[s1/(1-s1)]	

4.4. Synthesis

In both empirical analyses presented above we have observed the same pattern of intriguing results, viz. the existence of spatial flow models that can explain only a

[8] Buratto (1999) applied discrete choice models to the same data set and found that the share of chemical products shipped by road is price-inelastic (rigid, η_{sp}=0).

relatively small part of the total variation in the dependent variable but that still appear to retain a high predictive power. This apparent paradox calls for a closer examination of the mechanisms deployed during the empirical analysis. The statistical explanation of this phenomenon clearly emerges after inspection of the descriptive statistics of both data sets employed in the empirical analyses.

The predictive ability of the econometric models used in our analysis stems from the fact that, in general, the road flow accounts for a very large share of the total flow of goods between any two regions. Given the logistic requirements of the commodities shipped, the road system is almost a captive system. This means that road flows and total flows are, by definition, highly correlated. Moreover, the low coefficient of determination of the estimated models implies that the predicted share of freight shipped by road is almost a constant. Therefore, the predicted flow of freight shipped by road can almost directly be approximated by multiplying the total flow by the predicted share of freight shipped by road, and this is by definition highly correlated with the actual freight flow shipped by road. Thus, we may conclude that the high correlation between predicted and actual flows does not necessarily derive from the goodness-of-fit of the estimated models, but rather from structural patterns incorporated in the databases.

5. Conclusion

In this paper we have explored how the type of data stemming from spatial interaction models can be used to investigate how the flow of freight transport is distributed across competing transport modes. In the process, we have encountered various risks to which the researcher is exposed during the empirical analysis. The first risk is to draw too much of an analogy between freight transport data and passenger transport data. To apply the discrete choice models to freight transport data requires the explicit modelling of the interdependence between the decision making regarding the individual tonnes of freight and the other tonnes of freight belonging to the same shipment.

The second type of risk concerns the skewed and biased structural distribution of the flows of commodities over various transport modes. In this case, the total flow and the flow of commodities through the mostly used mode of transport are almost necessarily correlated. The correlation between flows implies that the predicted and

the observed flow of commodities are highly correlated by definition. This is especially true if the model does not fit the data very well. This phenomenon calls for both a thorough examination of the data and a close analysis of the residuals obtained from the estimated models. As a matter of fact, the estimated parameters and the results obtained can be trusted only after a close scrutiny of the model specification. Finally, a caveat is in order. Our analysis does not say that discrete choice models cannot be used in the analysis of spatial freight flows. Rather, we point to the fact that the knowledge of agents' behaviour in the carriers' market is of foremost importance, because it affects the specification of the econometric model.

Acknowledgments The authors are indebted to Philippe Tardieu (NEA Transport Research and Training, P.O. Box 1969, 2280 DZ Rijswijk, The Netherlands; http://www.nea.nl) for kindly providing the data and to Cees Gorter for valuable comments on an early version of the paper. The paper has substantially benefited from constructive comments from two anonymous referees.

References

Ben-Akiva M., Lerman S.R. (1985) *Discrete choice analysis: theory and application to travel demand*, MIT Press, London.

Buratto F. (1999) "I modelli di scelta discreta: analisi empiriche con riferimento alla domanda di trasporto merci" (Discrete choice models: empirical analysis with reference to the demand of freight transport), Bologna University, unpublished, in Italian.

Cramer J.S. (1991) *The logit model for economists*, Edward Arnold.

Cramer J.S. (1994) *Econometric applications of maximum likelihood methods*, Cambridge University Press, New York.

Cramer J.S. (1996) "Predictive performance in the binary logit model" mimeo, Tinbergen Institute, Amsterdam.

De Dios Ortúzar J., Willumsen L.G. (1990) *Modelling transport*, John Wiley & Sons, London.

Emmerink R.H.M. (1996) *Information and pricing in road transport*, Tinbergen Institute Research Series 123, Thesis Publishers, Amsterdam.

Green W. (2000) *Econometric analysis* Prentice Hall, Englewood Cliffs, New Jersey.

Gujarati D. N. (1995) *Basic Econometrics*, McGraw-Hill, Singapore.

Hurley W.J., Petersen E.R. (1996)a "Why regulate prices in freight transport markets" in Bianco L., Toth, P. (eds) *Advanced methods in transport analysis*, Springer, Heidelberg.

Hurley W.J., Petersen E.R. (1996)b "Optimal freight transport pricing and the freight network Equilibrium problem" in Bianco L., Toth, P. (eds) *Advanced methods in transport analysis*, Springer, Heidelberg.

Lee M. (1997) *Methods of moments and semiparametric econometrics for limited dependent variable models*, Cambridge University Press, New York.

Maddala G.S. (1985) *Limited dependent and qualitative variables in econometrics* Cambridge University Press, New York.

NCHRP (1997) "A guidebook for forecasting freight transportation demand" NCHRP Report 3888, Transportation Research Board, National Academy Press, Washington.

Nijkamp P., Reggiani A. (1989) "Spatial interaction and Input-Output models: a dynamic stochastic multiobjective framework" in Miller R.E., Polenske K.R., Rose A.Z. (1989) *Frontiers of Input-Output Analysis*, pp. 193 – 205, Oxford University Press, Oxford.

Nijkamp P., Reggiani A. (1998) *The economics of complex spatial systems*, Elsevier, Amsterdam.

Pregibon D. (1980) "Goodness of link tests for generalized linear models" *Applied Statistics*,19,15-24.

Reggiani A. (ed.) (1998) *Accessibility, trade, and locational behaviour*, Ashgate Aldershot, London.

Rietveld P., Nijkamp P. (2000) "Transport and regional development" in Polak J., Heertje A. (eds.) *Analytical transport economics*, pp. 208 – 234, Edward Elgar, Cheltenham.

APPENDIX 1: The Regional Classification Used